Cloning and characterization of mariner-like elements in the soybean aphid, *Aphis glycines* Matsumura

O. Mittapalli¹*, L. Rivera-Vega¹, B. Bhandary¹, M.A. Bautista¹, P. Mamidala¹, A.P. Michel¹, R.H. Shukle² and M.A.R. Mian³

¹Department of Entomology, The Ohio State University/Ohio Agricultural and Research Development Center, 1680 Madison Avenue, Wooster, OH 44691, USA: ²USDA-ARS and Department of Entomology, Purdue University, West Lafayette, IN 47907, USA: ³USDA-ARS and Department of Horticulture and Crop Science, The Ohio State University, 1680 Madison Avenue, Wooster, OH 44691, USA

Abstract

Soybean aphid, *Aphis glycines* Matsumura (Hemiptera: Aphididae), is currently the most important insect pest of soybean (*Glycine max* (L.) Merr.) in the United States and causes significant economic damage worldwide, but little is known about the aphid at the molecular level. Mariner-like transposable elements (MLEs) are ubiquitous within the genomes of arthropods and various other invertebrates. In this study, we report the cloning of MLEs from the soybean aphid genome using degenerate PCR primers designed to amplify conserved regions of mariner transposases. Two of the ten sequenced clones (designated as Agmar1 and Agmar2) contained partial but continuous open reading frames, which shared high levels of homology at the protein level with other mariner transposases from insects and other taxa. Phylogenetic analysis revealed Agmar1 to group within the irritans subfamily of MLEs and Agmar2 within the mellifera subfamily. Southern blot analysis and quantitative PCR analysis indicated a low copy number for Agmar1-like elements within the soybean aphid genome. These results suggest the presence of at least two different putative mariner-like transposases encoded by the soybean aphid genome. Both Agmar1 and Agmar2 could play influential roles in the architecture of the soybean aphid genome. Transposable elements are also thought to potentially mediate resistance in insects through changes in gene amplification and mutations in coding sequences. Finally, Agmar1 and Agmar2 may represent useful genetic tools and provide insights on *A. glycines* adaptation.

Keywords: Agmar1, Agmar2, mariner-like element, soybean aphid, transposable element

(Accepted 4 April 2011)
Introduction

The soybean aphid, *Aphis glycines* Matsumura (Hemiptera: Aphididae), is the most important insect pest of Soybean (*Glycine max* (L.) Merr.). Soybean aphids damage by feeding directly on plants and as a vector of various viruses (Wu et al., 2004). The United States is the world’s largest producer of soybean (91.4 million metric tons in the year 2009: FAO Statistical Database, 2010), but production in 2000 was threatened by the accidental introduction of the soybean aphid into North America (Ragsdale et al., 2007). An estimated 80% of the North American soybean crop is considered at risk (Vennette & Ragsdale, 2004).

Insecticide use is the most effective and practical method for soybean aphid control, but the environmental hazards, particularly on natural enemies, has necessitated research in host-plant resistance to the soybean aphid. A number of soybean accessions (plant introductions/collections from other countries) with resistance to the soybean aphid have been identified (Hill et al., 2004; Mensah et al., 2005; Mian et al., 2008a; Zhang et al., 2009), and the two single dominant genes, *Rag*1 (Hill et al., 2004) and *Rag*2 (Mian et al., 2008b), have been mapped and transferred into commercial soybean cultivars. Commercial cultivars with *Rag*1 were made available in 2010. The resistance by these genes in soybean is thought to operate via antibiotic, wherein aphids are killed as early nymphs. Before the widespread use of resistant cultivars, however, virulent soybean aphid populations have been recently discovered, termed as soybean aphid biotypes (Kim et al., 2008; Hill et al., 2010). At least three biotypes have been formally described, and the existence of others has been noted in field observations (A. Michel, personal communication). Biotype evolution has occurred through either genetic selection or recent changes at the genetic level. Answering this question depends on molecular resources which are currently lacking for this devastating insect pest.

Transposable elements belonging to the *mariner* family are thought to be the most widely distributed genetic elements in nature. These mobile elements are integral components in the genomes of various taxa, including bacteria, nematodes, insects, plants, fish and mammals (Hartl, 1989; Robertson, 1993; Robertson & Lampe, 1995a; Jarvik & Lark, 1998; Izsvák et al., 1999; Leroy et al., 2000; Schneider et al., 2000; Liu et al., 2007). *Mariner*-like elements (MLEs) are members of the *mariner*: *Tr1* transposon superfamily. The first *mariner* element was identified and characterized by Jacobson et al. (1986) in the white-peach mutant of *Drosophila mauritiana* Tsacas and David (Diptera: Drosophilidae). These functional elements encode a single protein, a transposase, which allow them to spontaneously incorporate into genomes via a cut-and-paste mechanism. Other characteristic features of MLEs include the presence of a short inverted terminal repeat and the requirement of a DNA intermediate for transposition. In view of their unique characteristics at the genome level across insect species, MLEs have been proposed to serve as valuable research tools, such as vectors for germline transformation, isolation and mapping of genes and transposon-based sequencing (Cooley et al., 1988; Strausbaugh et al., 1990; Maruyama & Hartl, 1991).

In general, two forms of MLEs exist: active and inactive, also termed autonomous and nonautonomous elements (Medhora et al., 1991). Inactive elements are simple derivatives of active forms with in-frame deletions that result in the loss of essential sequences required for transposition but retain other functional features. Thus, they are able to transpose only in the presence of active elements, which are sometimes referred to as mutator factors (Medhora et al., 1991). Many of the *mariner* elements are assumed to be inactive, and only four autonomous elements have been reported, including *Mos1* from *D. mauritiana* (Medhora et al., 1991); *Famar1* from the European earwig, *Forficula auricularia* (Barry et al., 2004); *Himar1* from the horn fly, *Haematobia irritans* (Bryan et al., 1987; Robertson & Lampe, 1995a); and *Mboumar-9* from the ant, *Messor bouvieri* (Muñoz-Lopez et al., 2008).

Given that transposable elements could serve as genetic tools and/or potentially mediate resistance in insects through changes in gene amplification and mutations in coding sequences (Li et al., 2007), we were interested in identifying MLEs in the soybean aphid genome. To our knowledge, this study would represent the first report of such functional elements and their implications in the soybean aphid genome.

Materials and methods

Experimental insect

The ‘Illinois’ (IL) colony samples were collected from a colony established in 2000 (Hill et al., 2004). The ‘Ohio’ (OH) colony was established from field collected aphids in Wooster, OH in the summer of 2005 (Kim et al., 2008). All colonies were established from a large number of randomly collected individuals from soybean. Individuals from each colony used in the current study were randomly collected from multiple soybean plants; only one aphid per plant was collected from randomly selected soybean plants to limit sampling of clones. The aphid colonies were maintained on seedlings of cultivar ‘Williams 82’ in growth chambers at temperatures between 22°C and 24°C with a photosynthetically active radiation of 330 µmol·m⁻²·s⁻¹ for 15h daily and 60–70% relative humidity (RH). DNA samples were stored at ~80°C in the lead author’s laboratory to serve as vouchers.

Homology-based cloning

Genomic DNA from 50 *A. glycines* individuals, including adults and nymphs, was isolated using the E.Z.N.A® Tissue DNA Kit from Omega Biotek, Inc (Doraville, GA, USA) following the manufacturers protocol. DNA isolated using the same protocol from 50 adult orange blossom wheat midge, *Sitodiplosis mosellana* (known to carry MLEs in the genome: Mittapalli et al., 2006), was used as the positive control in PCR reactions. DNA-negative controls were represented by water as the template in PCR reactions. Degenerate PCR primers designed to amplify regions of conserved amino acid residues in the middle of the *mariner* transposases as per Robertson (1993) with minor modifications were used (Mittapalli et al., 2006). The sequences of the degenerate primers are: MAR-159F, 5′-TNCNDKNGAYGARAMNTGG-3′ and MAR-249R, 5′-RYRTGNNGNSNCRTTRTC-3′, where D = A + G + T, K = G + T, M = A + C, N = A + C + G + T, R = A + G, S = C + G and Y = C + T. PCR cycling included: 94°C for 1 min, 55°C for 1 min and 72°C for 1 min for a total of 40 cycles, 72°C for 5 min, 4°C hold. Post amplification, 5 µl of each PCR reaction was analyzed on a 1% agarose gel. Four microliters of fresh PCR product was used to clone the amplified fragments directly.
into the PCR® 4-TOPO® vector included in a TOPO TA cloning® for sequencing kit (Invitrogen, Carlsbad, CA, USA) following the manufacturers protocol.

Sequence analysis

Ten independent clones (Agmar1–10) were sequenced using M13 sequencing primers over both strands at the Molecular and Cellular Imaging Center (MCIC) located at the Ohio Agricultural and Research Development Center (OARDC). Sequence analyses, including similarity and annotation searches, were done using different BLAST programs (Altschul et al., 1990) on the National Center for Biotechnology Information (Bethesda, MD, USA), web site (http://www.ncbi.nlm.nih.gov/). The deduced amino acid sequences for the soybean aphid MLEs were obtained using the ‘Translate tool’ (http://au.expasy.org/tools/dna.html). Multiple sequence alignment of the soybean aphid MLEs with MLE sequences from other insect species was performed using Multiple Sequence Comparison by Log Expectation (MUSCLE: Edgar, 2004).

Phylogenetic analysis

In order to reveal the relationship of the soybean aphid MLE sequences to other mariner transposases, a phylogenetic analysis was performed. A phylogram was constructed using the deduced amino acid sequences of the soybean aphid MLEs and different mariner transposase subfamilies, including mauritiana, irritans, capitata, mellifera and cecropia. The peptide sequences were aligned with the ClustalX program, version 8.1 with 11 updates (Thompson et al., 1997). The phylogram was calculated by the distance/neighbor-joining method using the software package PAUP* 4.0b10 (Swofford, 2002). The polypeptide sequence of a Tc1 transposable element transposase from Caenorhabditis elegans (accession no. X01005) represented the outgroup. Bootstrap values for the branches were obtained with 1000 replications.

Fig. 1. Mariner-like elements (MLEs) in the soybean aphid, Aphis glycines. (a) PCR amplification of MLEs in the soybean aphid using degenerate PCR primers for conserved regions of mariner transposases. Lanes are as follows: 1, Promega 1-kb ladder; 2, PCR fragments amplified using A. glycines DNA; 3, DNA-negative control PCR; 4, DNA-positive control using Orange blossom wheat midge (Sitodiplosis mosellana) DNA. (b) Partial MLE sequences obtained from A. glycines (Agmar1) without in-frame terminations ‘showing conserved domains of mariner transposases’. Highlighted with grey arrow-heads are two of the three aspartic acid (Asp, D) residues that constitute the D,D(34)D motif of the catalytic domain of Agmar1. Bold nucleotide sequences represent gene-specific primers for downstream analysis.
Southern blot analysis

The soybean aphid mariner element Agmar1 was used as the probe in Southern blot analysis. No EcoRI and XbaI restriction sites were identified within the nucleotide sequence obtained for Agmar1. Five micrograms (representing \( *50 \) aphid individuals) of genomic DNA from both biotypes of A. glycines ('OH' and 'IL') was digested individually by EcoRI restriction enzyme to generate DNA fragments with intact Agmar1 sequences. Agarose gel electrophoresis of the DNA fragments was conducted in a 0.8% agarose gel using 1X TAE buffer (0.04 mol l\(^{-1}\) Tris-acetate, 1 mmol l\(^{-1}\) EDTA, pH 8.0). DNA-negative lanes were included in the gel. The gel was treated and the DNA was transferred to Hybond N membrane (Amersham Pharmacia Biotech, Piscataway, NJ, USA) by standard blotting procedure (Sambrook & Russell, 2001). Hybridization was done at 65°C, and probes used were labeled using \( \alpha\)-32P]dCTP (Amersham Pharmacia Biotech) by oligonucleotide random priming (Feinberg & Vogelstein, 1983).

Copy number estimation of Agmar1 using quantitative PCR

Quantitative PCR has been previously reported in assessing the copy number of a gene (Applied Biosystems, Inc., Foster, CA, USA: Bubner & Baldwin, 2004; Moran et al., 2005; Solomon et al., 2008). The method involves relative quantification of the gene of interest vs. a reference gene, which is known to be present as a single copy in a haploid genome.

Approximately 20 mg of adult and nymph 'IL' aphids was used for DNA extraction following the Omega Bio-tek protocol described earlier. Two extractions were performed to represent three biological replicates in quantitative PCR (qPCR).

Primers for Agmar1 (Agmar1F2 5'-AGCGTTGTCGTGATGAAACA-3' and Agmar1R2 5'-ACTACACTCCTGAACAGGAAG-3') and for the reference gene RPL7 (RPL7F2 5'-CACACCGAATACCGCATGATAGATG-3' and RPL7R2 5'-CCCGAAGCCTATGTAAAGGAGTTC-3') were designed using Beacon Primers 7.5 (Premier Biosoft International, Palo Alto, CA, USA) based on partial genomic DNA and cDNA sequences of A. glycines, respectively.

The relative quantification method of CFX 96/C1000 BioRad Real Time System Software was used to approximate the copy number of Agmar1 in A. glycines genome. A standard curve, in which the reference gene ribosomal protein L7 (RPL7) is present in 100,000, 10,000, 1000 and 100 copies, was created following steps listed in an Applied Biosystems Inc. online manual. RPL7 is believed to be present as a single copy gene in the pea aphid (Acyrthosiphon pisum) genome and other genomes (Nikoh et al., 2010). Since A. glycines genome size is still unknown, the average size (ca. 500 Mb) of Aphis sp. (Finston et al., 1995) and A. pisum (TIAGC, 2010) was used for relative quantification.
used to calculate the amount of DNA needed to obtain the copy numbers for the standard curve. Spectrophotometric scan at 260nm was used to determine the amount of DNA needed.

PCR reactions were performed in a 20μl mixture. The mixture of one reaction contain 1X iQ SYBR Supermix (BioRad, Hercules, CA, USA), 5μM each of the primers, and 5μl genomic DNA (about 100ng). Nuclease-free water (Applied Biosystems, Inc.) was added to make the final volume. The thermal cycling conditions were: 95°C for 5min, followed by 40 cycles of 95°C for 10s and 60°C for 30s. A dissociation step was added to the protocol for a check on cross-contamination. Two experiments with three replications were conducted.

Results and discussion

PCR amplicons in the size range of 300bp were cloned from the genome of the soybean aphid (Aphis glycines) using degenerate primers designed to target conserved regions of mariner transposases (Robertson, 1993; Mittapalli et al., 2006). Similar-sized amplicons were also obtained from the orange blossom wheat midge (Sitodiplosis mosellana), which represented the positive control in this study (fig. 1a). No
amplicons were obtained for the negative control, wherein water was used as the template in PCR reactions.

A total of ten PCR clones were sequenced and evaluated as MLE transposases by BLAST searches. Two of the ten A. glycines clone sequences translated without in-frame terminations and represented unique sequences that contained conserved residues of mariner transposases (fig. 1b). By comparison, three uncharacterized nucleotide sequences categorized as ‘Mariner/Tc1’ family of transposons exist for the pea aphid, Acrithosiphon pisum Harris (Jurka et al., 2005). Both MLEs from A. glycines were designated by the authors as Agmar1 (clone Sm25-4) and Agmar2 (clone Sm38) and have the GenBank accessions numbers: GQ231493 and GQ231494, respectively. The other eight cloned sequences were not analyzed further because of multiple stops and in-frame shifts in their deduced protein sequences, suggesting the absence of any differences in copy number between biotypes. Data

Relative copy number

<table>
<thead>
<tr>
<th></th>
<th>RPL7</th>
<th>Agmar1</th>
<th>RPL7</th>
<th>Agmar1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expt 1</td>
<td><img src="image1" alt="Graph A" /></td>
<td><img src="image2" alt="Graph B" /></td>
<td><img src="image3" alt="Graph C" /></td>
<td><img src="image4" alt="Graph D" /></td>
</tr>
</tbody>
</table>

Fig. 4. Copy number estimation of Agmar1 in Aphis glycines. (a) Southern blot analysis of Aphis glycines genomic DNA to estimate the copy number of the mariner-like element, Agmar1. Lanes are as follows: 1, DNA ladder; 2, genomic DNA from ‘OH’ biotype of A. glycines digested with EcoR1; 3, genomic DNA from ‘IL’ biotype of A. glycines digested with EcoR1. (b) A histogram representing estimated copy number of Agmar1 relative to a single copy gene, RPL7, in two quantitative PCR experiments. Standard error of the mean of three technical replicates is shown.

Agmar1 with the other mariner transposase sequences are shown in fig. 2. Thus, the sequence analysis and molecular characterization provide first-hand evidence of the cloned PCR fragments from A. glycines genomic DNA are MLE transposase sequences.

Phylogenetic analysis was performed by constructing a neighbor-joining phylogram in accordance with previous studies (Robertson & MacLeod, 1993; Augé-Gouillou et al., 2000). The derived phylogram revealed five clades with good bootstrap support and the topology was in agreement with the expected groupings of the mariner transposase sequences into the subfamilies mauritiana, capitata, mellifera, cecropia and irritans (fig. 3). The two MLE transposase sequences from A. glycines grouped with the irritans (Agmar1) and mellifera (Agmar2) subfamilies. Specifically, Agmar1 grouped with a similar irritans transposase from Anopheles gambiae; whereas, Agmar2 grouped between Ceratitis capitata of the mellifera subfamily (fig. 3). These results confirm that the soybean aphid genome hosts at least two different subfamilies of mariner transposases. Given the wide distribution of similar elements in an array of arthropods (Robertson, 1993), it is plausible that the A. glycines genome hosts other subfamilies of mariner transposases than those reported in this study. This divergence of MLEs within a genome could result due to the loss in transposase-DNA interactions, i.e. minor base changes in the ITRs and coding regions of the transposase (Lampe et al., 2001).

Southern blot analysis indicated a low copy number for Agmar1 within the soybean aphid genomes of both biotypes including ‘IL’ and ‘OH’ (fig. 4a). There did not appear to be any differences in copy number between biotypes. Data
obtained via quantitative PCR (qPCR) further supported the Southern analysis, which also revealed a low copy number for Agmar 1 (fig. 4b). Agmar 2 was also estimated via qPCR to be low copy number in the genome of A. glycines (data not shown).

In insects, the copy numbers of MLEs per genome can be categorized into low, moderate, high and very high (Robertson & Lampe, 1995b). For example, 2–3 copies of MLEs have been estimated within the genomes of Drosophila ananassae (Robertson & Lampe, 1995a) and D. sechellia (Capy et al., 1992); 20–30 in Anopheles gambiae (Robertson & Lampe, 1995b), D. mauritiana (Maruyama & Hartl, 1991), Magetiola destructor (Shukle & Russell, 1995), Sitophilus mosellana (Mittapalli et al., 2006); 1000 in the moth Hyalophora cecropia (Lidholm et al., 1991; Robertson, 1993) and up to 17,000 copies in the horn fly, H. irritans (Robertson & Lampe, 1995a). Although the genome size of A. glycines is likely close to the size of other Aphid sp., including A. pismum (Finston et al., 1995; TIAGC, 2010), to state an exact copy number for the two retrieved MLEs within the A. glycines genome is still uncertain.

Although the function, evolution and transmission of MLEs is not fully known, it is thought that a single copy of these elements initially invades the genome of an organism, perhaps within a virus (Hartl et al., 1997), duplicates through transposition and finally becomes abundant in the population. The most common mechanism for such elements that transpose requiring a DNA intermediate has been proposed to be via horizontal transfer within members of the same order (Maruyama & Hartl, 1991; Simmons, 1992) and/or between members belonging to different insect orders (Robertson & Lampe, 1995b). The low copy numbers of Agmar1 within the soybean aphid genome suggest the transfer event(s) were fairly recent either being interspecific or intergeneric.

Some intriguing attributes have been assigned to the presence of transposable elements within insect systems. The mobile nature of transposable elements in insects could lead to the presence of transposable elements within insect systems. The mobile nature of transposable elements in insects could be low copy number in the genome of A. glycines (data not shown). For this to be achieved, full length sequences need to be obtained, and subsequent in vitro genetic analysis for transposition such as the ‘hop’ assays developed for Mbo2mar-9 (Muñoz-Lopez et al., 2008) need to be performed.

Acknowledgements

We thank Dr Esther van der Knaap’s laboratory for help with Southern blot analysis. Help with sequencing at the Molecular and Cellular Imaging Center (MCIC, Wooster, OH) is much appreciated. We also thank Curt Hill (University of Illinois) for sending the Illinois biotype aphids. The authors appreciate the comments/suggestions provided by the reviewers. Funding for this project was provided by State and Federal funds appropriated to the Ohio Agricultural Research and Development Center (OARDC, Wooster, OH), The Ohio State University and by the Ohio Soybean Council (OSC 08-2-08).

References


